



因素空间理论下的因果概率推理分类算法研究

曾繁慧, 胡光闪, 孙慧, 汪培庄

引用本文:

曾繁慧, 胡光闪, 孙慧, 汪培庄. 因素空间理论下的因果概率推理分类算法研究[J]. 智能系统学报, 2024, 19(4): 1042–1051.

ZENG Fanhui, HU Guangshan, SUN Hui, et al. A causal probabilistic inference classification algorithm based on factor space theory[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(4): 1042–1051.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202206004>

您可能感兴趣的其他文章

因素空间理论下基点分类算法研究

Base point classification algorithm based on factor space theory

智能系统学报. 2020, 15(3): 528–536 <https://dx.doi.org/10.11992/tis.201903031>

因素表示的信息空间与广义概率逻辑

Factorial information space and generalized probability logic

智能系统学报. 2019, 14(5): 843–852 <https://dx.doi.org/10.11992/tis.201810021>

概率粗糙集三支决策在线快速计算算法研究

Research on a fast online computing algorithm based on three-way decisions with probabilistic rough sets

智能系统学报. 2018, 13(5): 741–750 <https://dx.doi.org/10.11992/tis.201706047>

结合Copula理论与贝叶斯决策理论的分类算法

Classification algorithm based on Copula theory and Bayesian decision theory

智能系统学报. 2016, 11(1): 78–83 <https://dx.doi.org/10.11992/tis.201509011>

基于最小最大概率机的迁移学习分类算法

Transfer learning classification algorithms based on minimax probability machine

智能系统学报. 2016, 11(1): 84–92 <https://dx.doi.org/10.11992/tis.201505024>

SCADA安全因素神经元的云推理机研究与仿真

Study and simulation of the SCADA security factors neuron's cloud inference engine

智能系统学报. 2016, 11(5): 688–695 <https://dx.doi.org/10.11992/tis.201509020>

DOI: 10.11992/tis.202206004

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20231208.1652.002>

因素空间理论下的因果概率推理分类算法研究

曾繁慧^{1,2}, 胡光闪^{1,2}, 孙慧^{1,2}, 汪培庄^{1,2}

(1. 辽宁工程技术大学 理学院, 辽宁 阜新 123000; 2. 辽宁工程技术大学 智能工程与数学研究院, 辽宁 阜新 123000)

摘要: 机器学习方法与因果推理结合能极大地提升方法性能。为探究因果概率正逆向推理的分类效果, 基于因素空间理论下的因素概率论, 利用条件概率, 研究正向因素概率推理原理及模型并提出正向因果概率推理分类法 (forward causal probabilistic inference classification algorithm, FCPIC) 和简化条件的可取度分类法; 研究逆向因素概率推理原理及模型并结合贝叶斯网络提出逆向因果概率推理分类法 (reverse causal probabilistic inference classification algorithm, RCPIC)。将 3 个分类算法与 KNN (K-Nearest neighbor) 和 SVM (support vector machine) 算法进行实例对比验证, 研究结果表明: FCPIC 算法、可取度分类算法和 RCPIC 算法简单有效、具有可行性和实用性, 且可取度分类法和 RCPIC 算法性能优于 SVM 和 KNN 算法, FCPIC 算法对实际数据预测中必要类有查全需求的情况更优。研究结论丰富了因素空间的理论研究和应用价值。

关键词: 因素空间; 因果概率推理分类法; 可取度分类法; 贝叶斯网络; 因素概率论; 条件概率; 因果关系; 人工智能

中图分类号: TP18 文献标志码: A 文章编号: 1673-4785(2024)04-1042-10

中文引用格式: 曾繁慧, 胡光闪, 孙慧, 等. 因素空间理论下的因果概率推理分类算法研究 [J]. 智能系统学报, 2024, 19(4): 1042-1051.

英文引用格式: ZENG Fanhui, HU Guangshan, SUN Hui, et al. A causal probabilistic inference classification algorithm based on factor space theory[J]. CAAI transactions on intelligent systems, 2024, 19(4): 1042-1051.

A causal probabilistic inference classification algorithm based on factor space theory

ZENG Fanhui^{1,2}, HU Guangshan^{1,2}, SUN Hui^{1,2}, WANG Peizhuang^{1,2}

(1. College of Science, Liaoning Technical University, Fuxin 123000, China; 2. Institute of Intelligence Engineering and Mathematics, Liaoning Technical University, Fuxin 123000, China)

Abstract: The integration of machine learning techniques with causal reasoning can significantly enhance method performance. To investigate the classification effect of positive and reverse causal probability inferences, we rely our study on factor probability theory under factor space theory. Using conditional probability, we examined the principles and model of positive-factor probabilistic reasoning. This led to the proposal of the forward causal probabilistic inference classification algorithm (FCPIC) and a desirability classification method of simplified conditions. We also explored the principles and model of inverse factor probabilistic inference, which resulted in the proposal of the reverse causal probabilistic inference classification algorithm (RCPIC) along with a Bayesian network. The three classification algorithms were compared with the K-nearest neighbor (KNN) and support vector machine (SVM) algorithms. The results demonstrate that the FCPIC algorithm, the desirability classification algorithm, and the RCPIC algorithm are simple, effective, feasible, and practical. The performance of the desirability classification method and RCPIC algorithm surpasses those of both SVM and KNN. Additionally, the FCPIC algorithm is better when dealing with cases where the necessary classes in actual data prediction have full demand. These research findings contribute to the theoretical research and application value of factor space.

Keywords: factor space; causal probabilistic inference taxonomy; desirability taxonomy; Bayesian network; factor probability theory; conditional probability; causality; artificial intelligence

收稿日期: 2022-06-04. 网络出版日期: 2023-12-09.

基金项目: 辽宁省教育厅资助项目 (JYTQN2023210, LJK-ZZ20220047); 阜新市社会科学课题 (2023Fslx154, 2023Fslx017).

通信作者: 曾繁慧. E-mail: 597873883@qq.com.

1982 年, 汪培庄等^[1]提出了因素空间, 用来剖析事物的因果联系, 创建了因素空间理论^[2], 是国际智能数学最早同时问世的三个学派之一。相

关学者扩展了因素空间理论,并作了应用研究^[3-14],其中因素之间的因果关系分析^[5-6]为人工智能、数据挖掘提供了重要工具,例如,刘海涛等^[8]为因果分析提出了推理模型;曾繁慧等^[9]用因果推理算法与云模型相结合得到连续变量的决策树算法,这些研究都体现了一个重要的思想:背景关系决定因果推理。

因果推理在机器人技术研究中发挥了重要作用^[12],因果推理分为从因到果和从果到因两个方面,其中,从因到果是一种正向思维,从果到因则是一种逆向思维。如果把“推理”二字泛化,从因到果可以叫做正向推理,从果到因也可以叫做逆向推理。正向推理是从事物的表面出发,逆向推理是智能问题求解的普遍形式。智能问题是如何行动以达到目标,目标就是所希望达到的结果,例如,要使身体强健,首先就要反问:怎样才能使身体强健?身体强健是要达到的结果。在因素空间中,正向因果分析的条件因素是因,结果因素是果,而逆向因果分析是从果倒过来求因,也就是说,此时的结果因素是“因”,条件因素是“果”。

正向思维更专注表面的东西,追求的是短期的解决方案,而逆向思维可能更接近问题的本质,以找到恰当的方案,从长期来看会更有效果,所以正向和逆向推理思维都具有研究意义。Reichenbach等^[15]试图在概率术语中定义因果关系。Hájek等^[16]在概率因果关系一章中使用因果贝叶斯网表示因果关系,加强了概率到因果关系的推断。Sprenger^[17]给出了因果强度作为差异制造的概率理论的公理化基础,Faghihi等^[18]认为概率模糊逻辑是在机器中实现因果关系的更好选择,并将因果关系整合到机器中的理论框架,这是创建可解释人工智能的关键一步。本文基于因素空间理论下的因素概率论,利用条件概率,首先研究了正向因素概率推理模型,结合贝叶斯网络^[19],得到正向因果概率推理分类法(forward causal probabilistic inference classification algorithm, FCPIC);并根据正向因果推理,提出相的可取度,得到可取度分类法;其次根据逆向思维,提出逆向因素概率推理原理,建立逆向因素概率推理模型,结合贝叶斯网络,提出逆向因果概率推理分类法;最后进行实例验证。

1 预备知识

1.1 可测因素概率场

因素是描述万事万物的一把钥匙,可以理解为广义的基因。基因是生物的质根,就像每个基

因串出一列生物形态一样,每个因素串出事物的一系列相或属性。例如颜色这个因素就可以串出红、黄、蓝等性状(相)。从数学观点来看,因素是一种特殊的映射,把对象映射到其相上。因素空间^[1-2]相关基础理论如下:

定义1 设非空集合 D 为论域, I 为论域上一类性状的集合,从论域 D 到 I 的一个满射 f :

$$f: D \rightarrow I$$

称为 D 上一个因素,论域 D 就是一群实体或对象的集合,叫做 f 的定义域。

定义2 (性状空间)记

$$I = I(f) = \{a_1, a_2, \dots, a_k\}$$

叫做 f 的性状空间(或相域),其中各相彼此不同,或记作 $X(f)$ 。将 f_1, f_2, \dots, f_n 的性状空间 $I(f)$ 记为 $I(f_j) = \{a_1^j, a_2^j, \dots, a_k^j\} (j = 1, 2, \dots, n)$ 。

相域 $I(f) = \{a_1, a_2, \dots, a_k\}$,记

$$[a_t] = dD[f(d) = a_t, t = 1, 2, \dots, k]$$

叫做相的反馈外延。

定义3 给定因素空间 (D, F) ,如果 $f_1, f_2, \dots, f_n \in F$ 两两的析取为零因素, $f_i \vee f_j = 0 (1 \leq i \neq j \leq n)$,则

$$I(f_1 \wedge f_2 \wedge \dots \wedge f_n) = I(f_1) \times I(f_2) \times \dots \times I(f_n)$$

称 (D, F) 是一个简单因素空间。

定义4 (背景关系)给定因素空间 (D, F) ,记 $I = I(f_1) \times I(f_2) \times \dots \times I(f_n)$, $R = c = (c_1, c_2, \dots, c_n) \in I | \exists d \in D; f_1(d) = c_1, f_2(d) = c_2, \dots, f_n(d) = c_n$

称 R 为 f_1, f_2, \dots, f_n 的背景关系或背景集。其中,因素 f_1, f_2, \dots, f_n 互不相关。

定义5 考虑因素 $F = \{f_1, f_2, \dots, f_n\}$,性状空间:

$$I(F) = I(f_1) \times I(f_2) \times \dots \times I(f_n)$$

若对任意 $c = \{c_1, c_2, \dots, c_n\} \in I(F)$ 都有 $[c] = F^{-1}(c) = \{d \in D | F(d) = c\}$,则称 F 是可测因素。即 $f_1, f_2, \dots, f_n \in F$ 可测,将 $I(F)$ 记为 I_F 。

因素空间可测是解决不确定背景关系的基础。因素可测时,可以将每个因素都理解为随机变量,将每个因素中的各相理解为随机事件。例如,射击靶子的结果 ξ 具有相域 $I(\xi) = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ 。其中,因素为随机变量 ξ ,该因素的11个相为随机事件,即环数。

定义6 称 $\psi = (D, I_F)$ 为因素空间上所定义的可测空间,简称可测因素空间,对于定义于其上的概率测度 p ,称 (D, I_F, p) 为因素空间上所定义的概率场,简称可测因素概率场^[20-21]。

1.2 因果分析表

一个以对象为行(即论域),以条件因素和结果因素(最后一列)为列,表中第 i 行第 j 列元素是第 i 对象在第 j 因素下的状态,称这个表为因果

分析表,表头记作

$$(D, f_1, f_2, \dots, f_n; g)$$

因果分析表的每一行是一个对象在因素空间中的坐标;一张因素分析表就是由有限个对象在因果空间中所形成的一个背景样本,其样式见表1。

表1 因果分析表
Table 1 Causal analysis table

D	$F \rightarrow g$				
	f_1	f_2	\dots	f_n	g
d_1	$f_1(d_1)$	$f_2(d_1)$	\dots	$f_n(d_1)$	$g(d_1)$
d_2	$f_1(d_2)$	$f_2(d_2)$	\dots	$f_n(d_2)$	$g(d_2)$
\vdots	\vdots	\vdots	\dots	\vdots	\vdots
d_m	$f_1(d_m)$	$f_2(d_m)$	\dots	$f_n(d_m)$	$g(d_m)$

1.3 贝叶斯网络

贝叶斯网络是一种概率图模型,是一种模拟人类推理过程中因果关系的不确定性处理模型,其网络拓扑结构是一个有向无环图。

贝叶斯网络的有向无环图中:节点(圈)表示随机变量;箭头表示条件依赖,连接有因果关系、非条件独立的变量或命题。若两个节点间以单箭头连接在一起,其中一个节点是“因”,另一个节点是“果”,则两节点产生一个条件概率值。

例如,假设节点 A 直接影响到节点 B ,即 $A \rightarrow B$,则用从 A 指向 B 的箭头建立结点 A 到结点 B 的有向弧 (A, B) ,权值(连接强度)用条件概率 $p(B|A)$ 来表示,如图1所示。

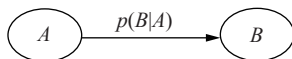


图1 AB 贝叶斯网络

Fig. 1 AB Bayesian network

根据图1可得 $p(AB)=p(B|A)p(A)$ 。

简言之,将研究系统中所涉及的随机变量,根据是否条件独立绘制在一个有向图中,就形成了贝叶斯网络。

2 正向因果概率推理分类法

为了利于分析,将表1化简得到因果分析简表,见表2。

表2 因果分析简表
Table 2 Causal analysis table

D	$F \rightarrow g$				
	f_1	f_2	\dots	f_n	g
d	c_1	c_2	\dots	c_n	g_i

由表2可得:

条件因素 $F=\{f_1, f_2, \dots, f_n\}$, $I(f_j)=\{a_1^j, a_2^j, \dots, a_k^j\} (j=1, 2, \dots, n)$ 。

结果因素 $g=\{g_1, g_2, \dots, g_s\}$, 当 $s=2$ 时,称为二分类问题,当 $s>2$ 时,则称为多分类问题。

对象 d (即样本)为一个 n 维向量,记 $d=[c_1, c_2, \dots, c_n]$, 其中 c_1, c_2, \dots, c_n 分别表示对象 d 在 f_1, f_2, \dots, f_n 下对应的相。如 $I(f_1)=\{a_1^1, a_2^1, \dots, a_k^1\}$, 则 $c_1=a_t^1 (t=1, 2, \dots, k)$ 。

2.1 正向因素概率推理

2.1.1 条件概率

设 A, B 为两个事件,且 $p(B)>0$,则在事件 B 发生的条件下,事件 A 发生的条件概率为

$$p(A|B) = \frac{p(AB)}{p(B)}$$

当 A 和 B 事件相互独立时, $p(AB)=p(A)p(B)$ 。条件概率包含了正向与逆向思维。当事件 A 为结果,事件 B 为条件时,则为正向分析,即正向推理;当事件 A 为条件,事件 B 为结果时,则为逆向分析,即逆向推理。

2.1.2 正向因素概率推理原理

推理有确定性推理和不确定性推理之分。例如,“若太阳升起,则天会亮”,这是一个推理句,它是确定性推理句,其真值(即正确率)是百分之百 $100\%=1$;“若同时掷两颗骰子,则两颗骰子点数之和大于2”,这是不确定性推理句,它多半是对的,但也有不对的时候。不确定性推理被概率论提升到科学的高度,因为这句话的正确率可以非常精确地表达成 $35/36!$

在因素概率论中,条件概率就是不确定性推理句的真值:

$$p(B|A) = t(AB) \quad (t \text{ 表示真值})$$

定义7 在因素空间 (D, F) 中,由已知的条件因素 f_1, f_2, \dots, f_n 寻找结果类别 g 的过程,即某对象由已知各因素 $f_j (j=1, 2, \dots, n)$ 所对应的相寻求结果类别的过程称为由因寻果。

2.1.3 正向因素概率推理模型

正向推理是有目标的求索。

给定事件列 a (a 为因素对应的相)和结果类 g_1, g_2, \dots, g_s , 已知条件概率 $p(g_i|a)$ 。

若 a 已经发生,要问 g 类事件中,哪个事件最有可能发生? 这就是正向因素概率推理。

定理1 给定事件列 a 和结果类 g_1, g_2, \dots, g_s , 已知结果 g 类事件中每个事件在 a 下的条件概率为 $p(g_i|a)$, 若

$$i^* = \operatorname{argmax} \{p(g_i|a) | i=1, 2, \dots, s\}$$

则事件 f_i 最有可能发生。

按照条件概率的定义,证明是显然的。

2.2 正向因果概率推理分类法——FCPIC

根据正向因素概率推理原理,利用正向因素

概率推理模型,下面研究正向因果概率推理分类法(FCPIC)。

假设1 因素空间 (D, F) 中因素 f_1, f_2, \dots, f_n 之间相互条件独立。

假设2 因素 f_1, f_2, \dots, f_n 对应的各相转换的数据均是离散型数据。

正向因素概率推理分类是由因寻果,即由已知条件寻找结果,也即某对象由已知各因素 f_j ($j=1, 2, \dots, n$) 所对应的相寻求结果类别。

1) 正向因果概率推理分类法。

在因素空间中,条件因素 f_1, f_2, \dots, f_n 与结果因素 g 对应的正向贝叶斯网络可由图2表示。

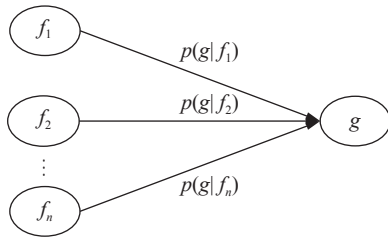


图2 条件和结果因素正向贝叶斯网络

Fig.2 Condition and outcome factor forward Bayesian network

当分析对象 d 时, $d=[c_1 \ c_2 \ \dots \ c_n]$, 即 $I(f_1)=c_1$, $I(f_2)=c_2, \dots, I(f_n)=c_n$ 有

$$p(g|I(f_1)) = p(g|I(f_1) = c_1)$$

$$p(g|I(f_2)) = p(g|I(f_2) = c_2)$$

⋮

$$p(g|I(f_n)) = p(g|I(f_n) = c_n)$$

为了简便,简记为

$$p(g|c_1), p(g|c_2), \dots, p(g|c_n) \quad (1)$$

由图2的正向贝叶斯网络可得

$$p(g, c_1, c_2, \dots, c_n) = p(c_1) \times p(c_2) \times \dots \times p(c_n) \times p(g|c_1) \times p(g|c_2) \times \dots \times p(g|c_n) \quad (2)$$

由正向推理模型和式(2)可得

$$p(g_i|c_1, c_2, \dots, c_n) = \frac{p(g_i, c_1, c_2, \dots, c_n)}{p(c_1, c_2, \dots, c_n)} = \frac{(p(c_1) \times p(c_2) \times \dots \times p(c_n) \times p(g_i|c_1) \times p(g_i|c_2) \times \dots \times p(g_i|c_n))}{p(c_1) \times p(c_2) \times \dots \times p(c_n)}$$

即

$$p(g_i|d) = \prod_{j=1}^n p(g_i|c_j), \quad i=1, 2, \dots, s \quad (3)$$

称式(3)是正向因果概率推理分类公式,表示对象 d 为结果 g_i 时发生的可能性。

式(3)中, c_j 表示对象 d 在第 j 个因素上的相取值; $p(g_i|c_j) = \frac{|D_{g_i, c_j}|}{|D_{c_j}|}$, 其中 $|D_{g_i, c_j}|$ 表示在因素 f_j 相值为 c_j 的训练集中类别为 g_i 的样本数量, $|D_{c_j}|$ 表示 D 中 f_j 相值为 c_j 的样本数量。

设 $N(d, g_i^*)$ 为对象 d 的最终判定类别,则正向因果概率推理分类公式为

$$N(d, g_i^*) = \operatorname{argmax}_{j=1}^n p(g_i|c_j), \quad i=1, 2, \dots, s \quad (4)$$

正向因果概率推理分类法是通过计算对象在不同结果类别下发生的可能性来达到预测结果分类的目的。

2) 正向因果概率推理分类公式的修正。

如果 $|D_{g_i, c_j}| = 0$, 即 c_j 为结果类 g_i 的数量为 0, 根据式(3), 不论其他因素的相如何, 对于这一结果类的对象发生概率值都为 0。为避免这种情况, 提出了解决办法, 对其进行“平滑”处理, 使用“拉普拉斯修正”法:

$$p(g_i|c_j) = \frac{|D_{g_i, c_j}| + 1}{|D_{c_j}| + |N_g|} \quad (5)$$

式中 $|N_g|$ 表示研究系统的结果类别个数。

利用修正的正向因果概率推理分类法解决例1问题。

例1 电风扇质量检测结果的因果分析见表3。

其中, f_1 为扇叶数量, f_2 为每分钟转速, f_3 为散热性能, g 为检测结果。

$$I(f_1) = \{\text{五扇叶, 七扇叶}\} = \{a_1^1, a_2^1\} = \{1, 2\};$$

$$I(f_2) = \{\text{高速, 中速, 低速}\} = \{a_1^2, a_2^2, a_3^2\} = \{1, 2, 3\};$$

$$I(f_3) = \{\text{好, 中, 差}\} = \{a_1^3, a_2^3, a_3^3\} = \{1, 2, 3\};$$

$$I(g) = \{\text{好, 差}\} = \{g_1, g_2\} = \{1, 2\}。$$

表3 电风扇质量因果分析表

Table 3 Electric fan quality causal analysis table

D	$F \rightarrow g$			
	f_1	f_2	f_3	g
d	c_1	c_2	c_3	g_i
d_1	1	3	1	2
d_2	1	2	3	2
d_3	2	1	3	2
d_4	2	1	1	1
d_5	2	2	1	1
d_6	2	3	2	2
d_7	2	1	2	1
d_8	1	1	1	1
d_9	2	3	1	2
d_{10}	1	2	2	2

利用修正的正向因果概率推理分类公式, 在 $j=1, i=1, 2$ 时, 即因素为 f_1 , 对 $g_1=1, g_2=2$ 的计算结果分别为

$$\textcircled{1} \ c_1 = 1, p(g_1|c_1) = \frac{|D_{g_1, c_1}| + 1}{|D_{c_1}| + |N_g|} = \frac{1 + 1}{4 + 2} = \frac{2}{6};$$

$$p(g_2|c_1) = \frac{|D_{g_2, c_1}| + 1}{|D_{c_1}| + |N_g|} = \frac{3 + 1}{4 + 2} = \frac{4}{6}。$$

$$\textcircled{2} c_1 = 2, p(g_1|c_1) = \frac{3+1}{6+2} = \frac{4}{8};$$

$$p(g_2|c_1) = \frac{3+1}{6+2} = \frac{4}{8}.$$

以此类推, 计算过程省略, 针对 d_1 :

$$N(d_1) = \left[\frac{2}{6} \times \frac{1}{5} \times \frac{4}{7}, \frac{4}{6} \times \frac{4}{5} \times \frac{3}{7} \right] = \left[\frac{4}{105}, \frac{8}{35} \right]$$

即 d_1 发生时, 结果为 g_2 的可能性更大, 所以将 d_1 判定为 g_2 类。

3 可取度分类法

当 $|D_{g_i, c_j}| = 0$ 时, 利用式(3)求得对象的这一结果类概率必为 0, 进而影响分类预测结果的准确性, 为了避免这种情况, 除了平滑处理外, 运用正向推理思想, 考虑各因素的各相对各结果类的影响程度作为分类标准, 给出相的可取度, 提出适用范围更广的可取度分类法。

3.1 相对结果因素的可取度

定义 8 (相可取度) 给定一个条件因素 f_j 及该因素所取的一个状态 a , 记 $[a] = \{d | I(f_j(d)) = a\}$, 对结果因素的状态 g_i , 使得 $[g_i] = \{d | I(g(d)) = g_i\}$ $|D_{g_i}[a](i=1, 2, \dots, s)$, 则称 $[a]$ 为 g_i 的一个单可取相。结果类为 g_i 的相 a 所占行数 h 与 g_i 所占的行数 $|D_{g_i}|$ 之比称为 a 对结果 g_i 的可取度, 记为 $q(a, g_i) = h/|D_{g_i}|$ 。

相可取度是由条件因素对结果因素的正向推理, 因此可称为由因寻果。

3.2 可取度分类方法

根据相可取度定义, 引申为对象可取度。

定义 9 对于一个对象 d , $d=(c_1, c_2, \dots, c_n)$, 结果因素的状态为 g_i 。将 c_1, c_2, \dots, c_n 对结果类 g_i 的可取度之和称为对象 d 对结果类 g_i 的可取度。

定理 2 给定事件列 a_1, a_2, \dots, a_k 和结果类 g_1, g_2, \dots, g_s , 已知 a_1, a_2, \dots, a_k 对结果 g 类事件的可取度 $q(a_q, g_i)$, 若

$$a_{q^*} = \arg\max \frac{|a_q \cap g_i|}{|D_{g_i}|}, \quad i=1, 2, \dots, s, q=1, 2, \dots, k \quad (6)$$

则事件 a_{q^*} 最有可能发生。

根据定理 2, 得到可取度分类公式为

$$q(d, g_i^*) = \arg\max \sum_{j=1}^n \frac{|c_j \cap g_i|}{|D_{g_i}|}, \quad i=1, 2, \dots, s, j=1, 2, \dots, n \quad (7)$$

式中: $|c_j \cap g_i|$ 表示相 c_j 为结果类 g_i 所占的行数, g_i^* 为对象 d 的最终判定类别。

可取度分类法是通过计算对象在不同结果类别下的可取度大小, 从而判别 d 的类别。

注: 可取度分类法不需要条件因素相互条件独立, 因此该算法适用范围更广。

例 2 利用可取度分类法解决例 1 问题。

首先计算所有条件因素的相可取度, 以因素 f_1 扇叶数量为例, 将论域 D 分为 2 类:

a_1^1 =五扇叶: $\{d_1, d_2, d_8, d_{10}\}$,

a_2^1 =七扇叶: $\{d_3, d_4, d_5, d_6, d_7, d_9\}$ 。

1) a_1^1 类中对象对应的结果有好有坏, 可继续划分:

五扇叶 $a_1^1 \rightarrow g_1: \{d_8\}$,

五扇叶 $a_1^1 \rightarrow g_2: \{d_1, d_2, d_{10}\}$ 。

相 a_1^1 五扇叶对结果 g_1 的可取度为 $q(a_1^1, g_1) = h/|D_{g_1}| = 1/4 (h=1)$;

相 a_1^1 五扇叶对结果 g_2 的可取度为 $q(a_1^1, g_2) = h/|D_{g_2}| = 3/6 (h=3)$ 。

2) $|D_{g_1}|$ 类中对象对应的结果有好有坏, 可继续划分:

七扇叶 $a_2^1 \rightarrow g_1: \{d_4, d_5, d_7\}$,

七扇叶 $a_2^1 \rightarrow g_2: \{d_3, d_6, d_9\}$ 。

相 a_2^1 七扇叶对结果 g_1 的可取度为 $q(a_2^1, g_1) = h/|D_{g_1}| = 3/4 (h=3)$;

相 a_2^1 七扇叶对结果 g_2 的可取度为 $q(a_2^1, g_2) = h/|D_{g_2}| = 3/6 (h=3)$ 。

以此类推, 计算过程省略, 针对 d_1 :

$$q(d_1) = \left[\frac{1}{4} + \frac{0}{4} + \frac{3}{4}, \frac{3}{6} + \frac{3}{6} + \frac{2}{6} \right] = \left[1, \frac{4}{3} \right]$$

即对象 d_1 对结果类 g_1 的可取度为 1, 对结果类 g_2 的可取度为 $4/3$, 因此对象 d_1 判别为 g_2 类。

4 逆向因果概率推理分类法

正向推理是无目标的求索, 目标锁定时, 可进行逆向推理, 因此本节探讨逆向因素概率推理原理、逆向因素概率推理模型以及逆向因果概率推理分类法。

4.1 逆向因素概率推理

2.1 节是由已知的条件 a 寻求结果 g , 下面要由已知的结果 g 返回寻找条件 a 。

4.1.1 逆向因素概率推理原理

给定事件列 a_1, a_2, \dots, a_k 以及结果列 g , 其中 a_1, a_2, \dots, a_k 是某因素的全相列。已知条件概率 $p(g|a_q) (q=1, 2, \dots, k)$, 现在要反过来问: 在 g 发生的条件下哪个事件最有可能发生?

若 $q^* = \arg\max \{p(g|a_q) | q=1, 2, \dots, k\}$, 则得出结论 a_{q^*} 最有可能发生。该结论不一定成立, 因为从 $p(g|a_{q^*})$ 最大推出 $p(a_{q^*}|g)$ 最大是没有根据的。例如, 一个班里 10 名女生, 其中 4 名数学得优, $p(g|a_1) = 2/5$, 有 20 名走读生, 有 5 名得优, $p(g|a_2) = 1/4$, $p(g|a_1)$ 高于 $p(g|a_2)$; 但全班共 8 名得

优, $p(a_1|g)=4/8$, $p(a_2|g)=5/8$ 。 $p(a_1|g)$ 低于 $p(a_2|g)$, 这就是一个反例。

定义 10 在因素空间 (D, F) 中, 由已知的结果类别 g 寻找条件因素 f_1, f_2, \dots, f_n 的过程, 即由已知结果类别寻求某对象各因素 $f_j (j=1, 2, \dots, n)$ 所对应的的相的过程称为由果寻因。

4.1.2 逆向因素概率推理模型

一般来讲, 最有可能发生的事件是谁? 给定事件列 a_1, a_2, \dots, a_k , 结果列 g , 已知 g 类事件的概率 $q^* = \operatorname{argmax} \{p(a_q|g)p(g)|q=1, 2, \dots, k\}$ 以及条件概率 $q^* = \operatorname{argmax} \{p(a_q|g)p(g)|q=1, 2, \dots, k\}$, 现在 g 已经发生, 最有可能发生的事件是谁?

根据条件概率公式, 可知

$$p(a_q|g) = \frac{p(g, a_q)}{p(g)} \quad (8)$$

定理 3 给定事件列 a_1, a_2, \dots, a_k , 结果列 g , 已知 g 类事件的概率 $q^* = \operatorname{argmax} \{p(a_q|g)p(g)|q=1, 2, \dots, k\}$ 以及条件概率 $p(a_q|g)$, g 发生时, 若

$$q^* = \operatorname{argmax} \{p(a_q|g)p(g)|q=1, 2, \dots, k\} \quad (9)$$

则最有可能发生的事件是 a_{q^*} 。

按照条件概率的定义, 证明是显然的。

4.2 逆向因果概率推理分类法——RCPIC

根据逆向因素概率推理原理, 利用逆向因素概率推理模型, 下面研究逆向因果概率推理分类法 RCPIC。

假设 1 因素空间 (D, F) 中因素 f_1, f_2, \dots, f_n 之间相互条件独立。

假设 2 因素 f_1, f_2, \dots, f_n 对应的各相转换的数据均是离散型数据。

逆向因素概率推理分类法是由果寻因, 即由已知结果寻找条件, 也即由结果类别寻求对象各因素 $f_j (j=1, 2, \dots, n)$ 所对应的相。

在因素空间中, 条件因素 f_1, f_2, \dots, f_n 与结果因素 g 对应的逆向贝叶斯网络可由图 3 表示。

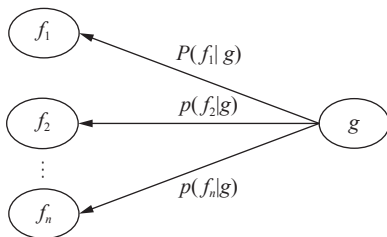


图 3 条件和结果因素逆向贝叶斯网络

Fig. 3 Condition and outcome factor inverse Bayesian network

根据表 2, 分析对象 d 时, 因 $d=[c_1 \ c_2 \ \dots \ c_n]$, 即 $I(f_1(d))=c_1, I(f_2(d))=c_2, \dots, I(f_n(d))=c_n$ 有

$$p(I(f_1)|g) = p(I(f_1)=c_1|g)$$

$$p(I(f_2)|g) = p(I(f_2)=c_2|g)$$

\vdots

$$p(I(f_n)|g) = p(I(f_n)=c_n|g)$$

为了简便, 简记为

$$p(c_1|g), p(c_2|g), \dots, p(c_n|g) \quad (10)$$

由图 3 的逆向贝叶斯网络可得

$$p(g, c_1, c_2, \dots, c_n) = p(g_i) \times p(c_1|g_i) \times p(c_2|g_i) \times \dots \times p(c_n|g_i) \quad (11)$$

由逆向推理模型和式 (11) 可得

$$p(c_1, c_2, \dots, c_n|g_i)p(g_i) = \frac{p(g_i, c_1, c_2, \dots, c_n)}{p(g_i)}p(g_i) = p(g_i) \times p(c_1|g_i) \times p(c_2|g_i) \times \dots \times p(c_n|g_i)$$

即

$$p(d|g_i)p(g_i) = p(g_i) \prod_{j=1}^n p(c_j|g_i), \quad i=1, 2, \dots, s \quad (12)$$

称式 (12) 是逆向因果概率推理分类公式, 为结果 g 发生时对象 d 发生的可能性。

式 (12) 中, c_j 表示对象 d 在第 j 个因素上的相取值。 $p(g_i) = \frac{|D_{g_i}|}{|D|}$, 其中 $|D|$ 为研究的论域中总对象数量, $|D_{g_i}|$ 为研究的论域中结果类别为 g_i 的对象数量。 $p(c_j|g_i) = \frac{|D_{c_j, g_i}|}{|D_{g_i}|}$, 其中 $|D_{c_j, g_i}|$ 表示在研究的论域中, 结果类别为 g_i 时, 因素 $f_j = c_j$ 的对象数量。

设 $T(d, g_i^*)$ 为对象 d 的最终判定类别, 则逆向因果概率推理分类法的表达式为

$$T(d, g_i^*) = \operatorname{argmax}_p(g_i) \prod_{j=1}^n p(c_j|g_i), \quad i=1, 2, \dots, s \quad (13)$$

逆向因果概率推理分类法是通过计算在不同结果类别下对象发生的可能性来达到预测结果分类的目的。

同理, 在类别为 g_i 时, 如果出现 c_j 的样本个数是 0, 即 $|D_{c_j, g_i}|=0$ 的情况, 根据式 (12) 不论其他因素相如何, 那么对于类别 g_i 的选择概率值都会是 0。为避免这种情况, 可进行“平滑”处理, 使用“拉普拉斯修正”法, 计算公式为

$$p(c_j|g_i) = \frac{|D_{c_j, g_i}| + 1}{|D_{g_i}| + |N_j|} \quad (14)$$

式中: $|N_j|$ 表示因素 f_j 的类别个数, 如 f_1 的相取值包括 1 和 2, 那么此时的 $|N_1|=2$ 。

例 3 利用逆向因果概率推理分类法解决例 1 问题。

$$p(g_1) = \frac{|D_{g_1}|}{|D|} = \frac{4}{10}, \quad p(g_2) = \frac{|D_{g_2}|}{|D|} = \frac{6}{10}$$

$i=1, j=1, 2, 3$ 时, 即 $g_1=1$ 时对各条件因素的计算结果分别为

$$\textcircled{1} \ c_1=1, p(c_1|g_1) = \frac{|D_{c_1, g_1}| + 1}{|D_{g_1}| + |N_1|} = \frac{1+1}{4+2} = \frac{2}{6}$$

$$c_1 = 2, p(c_1|g_1) = \frac{|D_{c_1, g_1}| + 1}{|D_{g_1}| + |N_1|} = \frac{3 + 1}{4 + 2} = \frac{4}{6}$$

$$\textcircled{2} c_2 = 1, p(c_2|g_1) = \frac{|D_{c_2, g_1}| + 1}{|D_{g_1}| + |N_2|} = \frac{3 + 1}{4 + 3} = \frac{4}{7}$$

$$c_2 = 2, p(c_2|g_1) = \frac{|D_{c_2, g_1}| + 1}{|D_{g_1}| + |N_2|} = \frac{1 + 1}{4 + 3} = \frac{2}{7}$$

$$c_2 = 3, p(c_2|g_1) = \frac{|D_{c_2, g_1}| + 1}{|D_{g_1}| + |N_2|} = \frac{0 + 1}{4 + 3} = \frac{1}{7}$$

$$\textcircled{3} c_3 = 1, p(c_3|g_1) = \frac{|D_{c_3, g_1}| + 1}{|D_{g_1}| + |N_3|} = \frac{4}{7}$$

$$c_3 = 2, p(c_3|g_1) = \frac{|D_{c_3, g_1}| + 1}{|D_{g_1}| + |N_3|} = \frac{2}{7}$$

$$c_3 = 3, p(c_3|g_1) = \frac{|D_{c_3, g_1}| + 1}{|D_{g_1}| + |N_3|} = \frac{1}{7}$$

以此类推, 计算过程省略, 针对 d_1 :

$$T(d_1) = \left[\frac{4}{10} \times \frac{2}{6} \times \frac{1}{7} \times \frac{4}{7}, \frac{6}{10} \times \frac{4}{8} \times \frac{4}{9} \times \frac{3}{9} \right] = \left[\frac{8}{735}, \frac{2}{45} \right]$$

即与结果 g_1 相比, 结果为 g_2 时, d_1 发生的可能更大, 所以将 d_1 判定为 g_2 类。

注: 本文提出 FCPIC、RCPIC、可取度分类算法的时间复杂度均为 $O(mns)$ 。

KNN 的算法复杂度为 $O(mn)$, SVM 的算法复杂度为 $O(m^2) \sim O(m^3)$ 。

其中 m 是训练集中对象个数, n 是条件因素的个数, s 是结果因素的相的个数。有 $O(mn) < O(mns) < O(m^2)$ 。

5 实例应用

为了验证 FCPIC、RCPIC、可取度分类算法在实际数据中的预测性能, 使用 UCI 数据库中的 3 个分类数据集并进行学习, 与 KNN 和 SVM 分类算法分别进行结果和时间复杂度对比。

各数据集信息如下:

Heart 数据集是诊断是否患有心脏病, 患病为正类, 不患病为负类。

Breast cancer 数据集是通过肿瘤的团块厚度、细胞大小的均匀性、细胞形状的均匀性等来判断是恶性还是良性, 恶性为正类, 良性为负类。

Australian 数据集是澳大利亚信贷批准, 批准为正类, 不批准为负类。

各数据集的具体结构见表 4。

表 4 数据集信息
Table 4 Dataset information

数据集	样本数	因素个数	结果类别	正类	负类
Heart	297	13	2	150	120
Breast cancer	683	9	2	239	444
Australian	690	14	2	383	307

分别对 5 种分类算法采用十折交叉验证, 得到准确率 Acc(accuracy)、精确率 Pre(precision)、召回率 Rec(recall)、 F_1 和 AUC, 其中精确率表示识别为正类的对象中正确的比例, 召回率表示正类对象被识别出的比例, 实验结果见表 5。

表 5 实验结果对比
Table 5 Comparison of experimental results

数据集	指标	FCPIC	RCPIC	可取度 分类法	KNN	SVM
Heart	Acc	0.8148	0.863	0.8704	0.8074	0.8481
	Pre	0.7632	0.8228	0.8365	0.8365	0.8717
	Rec	0.9667	0.8833	0.8833	0.8133	0.8533
	F_1	0.8528	0.8511	0.8581	0.8243	0.8620
	AUC	0.9144	0.9411	0.9467	0.9115	0.9272
Breast cancer	Acc	0.9809	0.9839	0.9252	0.9649	0.9589
	Pre	0.9757	0.9977	0.9081	0.9752	0.9052
	Rec	0.9707	0.9775	0.9865	0.9243	0.9875
	F_1	0.9727	0.9874	0.9454	0.9472	0.9438
	AUC	0.9882	0.9969	0.9872	0.9805	0.9879
Aust- ralian	Acc	0.8197	0.8706	0.8501	0.7483	0.8371
	Pre	0.7709	0.8795	0.8572	0.7204	0.8388
	Rec	0.9712	0.8238	0.8009	0.8926	0.8794
	F_1	0.8581	0.8489	0.8257	0.7968	0.8572
	AUC	0.9216	0.9326	0.9005	0.8417	0.8901

为直观地展示算法实验结果, 对比分析 5 种算法在不同数据集上的指标。

实验 1

对于 Heart 数据集, 各分类算法的准确率、精确率、召回率、 F_1 和 AUC 的指标对比关系见图 4。

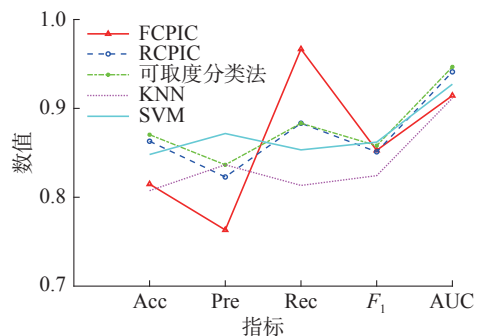


图 4 5 种算法在 Heart 数据集下的指标

Fig. 4 Five algorithm indicators under the Heart dataset

由图 4 和表 5 可知, 对于 Heart 数据集, RCPIC 算法和可取度分类法的准确率、召回率以及 AUC 都优于 KNN 和 SVM 算法; FCPIC 的性能优于 KNN 且召回率最高。

实验 2

对于 Breast cancer 数据集, 各分类算法的准

准确率、精确率、召回率、 F_1 和 AUC 的指标对比关系见图 5。

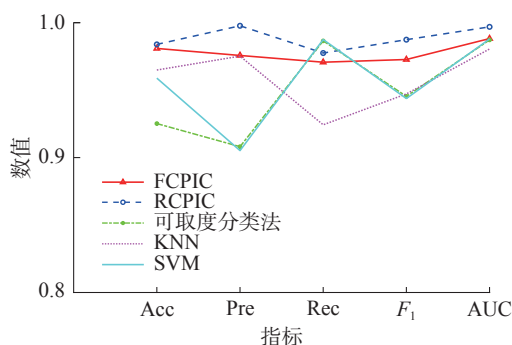


图 5 5 种算法在 Breast cancer 数据集下的指标

Fig. 5 Five algorithm indicators under the Breast cancer dataset

由图 5 和表 5 可知,对于 Breast cancer 数据集,FCPIC 和 RCPIC 算法的准确率、精确率、 F_1 以及 AUC 都优于 KNN 和 SNM 算法;可取度分法具有较高的召回率,其性能优于 KNN 算法。

实验 3

对于 Australian 数据集,各分类算法的准确率、精确率、召回率、 F_1 和 AUC 的指标对比关系见图 6,ROC 曲线见图 7。

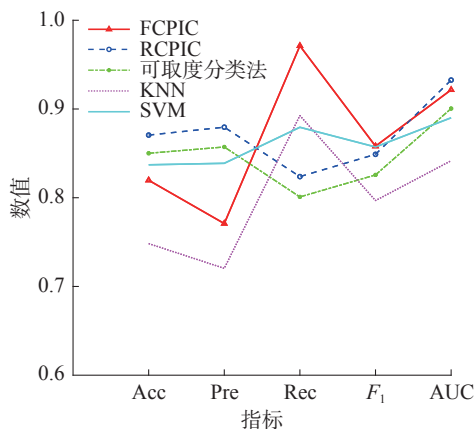


图 6 5 种算法在 Australian 数据集下的指标

Fig. 6 Five algorithm indicators under the Australian dataset

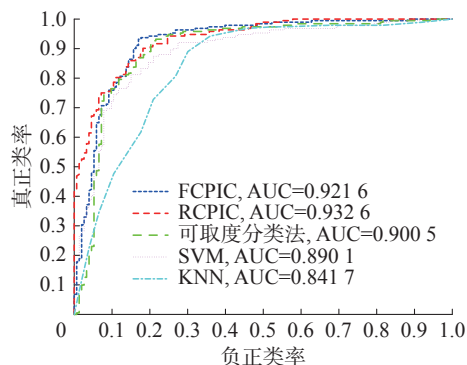


图 7 5 种算法在 Australian 数据集下的 ROC 曲线

Fig. 7 ROC curves of 5 algorithms on the Australian dataset

由图 6 和表 5 可知,对于 Australian 数据集,FCPIC、RCPIC 和可取度分类法的性能都优于 KNN 和 SNM 算法,且 FCPIC 的召回率最高,即更多地找到可以批准的信贷用户。

由图 7 和表 5 可知,RCPIC 算法的 ROC 曲线最凸,完全“包住”SVM 和 KNN,即随着算法划分对象为正类的个数增多时,划错的个数少于 SVM 和 KNN 算法;可取度分类法的 ROC 曲线没有完全“包住”SVM,但 AUC 值大于 SVM 和 KNN 算法,即性能优于 SVM 和 KNN 算法;FCPIC 算法的预测准确率低于 SVM 算法,但其左上方的 ROC 曲线更凸,性能更优。

综上分析,相较于 KNN 和 SVM 算法,FCPIC 具有较高的召回率,在查找患者等重要数据时,漏查率较低。RCPIC 算法的准确率和性能都优于 KNN 和 SVM 算法,4 个评价指标相对其他算法更平稳。可取度分类算法对不同数据集精确率或召回率较高,且性能也好。

算法复杂度对比

FCPIC、RCPIC 和可取度分类算法的时间复杂度低于 SVM 算法。

FCPIC、RCPIC 和可取度分类算法的时间复杂度稍高于 KNN 算法,即 $O(mns) > O(mn)$,但算法分类预测性能优于 KNN 算法。

6 结束语

本文基于因素空间理论下的因素概率论及条件概率,提出因果概率推理分类法,该算法能够清晰地阐述研究系统的构成,即由条件因素、结果因素及各因素的相构成。提出 3 种时间复杂度为 $O(mns)$ 的分类算法,与 SVM 和 KNN 算法进行比较,并用实例验证得到结论:

1)FCPIC 算法原理是由原因寻找结果的正向概率推理,与 KNN 和 SVM 算法相比具有较高的召回率,针对实际数据预测中必要类有查全需求的情况,优先选择该算法。

2)可取度分类法原理是计算原因对结果的影响程度的正向推理,可取度分类法的性能优于 KNN 算法,且与 SVM 算法相当。该算法无因素相互独立条件,适用范围更广。

3)RCPIC 算法原理是由结果寻找原因的逆向概率推理,其性能优于 KNN 和 SVM 算法,准确率、精确率、召回率和 F_1 4 个指标相对其他算法更平稳。

综上,本文在理论和实际层面分析了算法的有效性,在未来将进一步研究正逆向概率推理结合的情况。

参考文献:

- [1] 汪培庄, SUGENO M. 因素场与模糊集的背景结构 [J]. 模糊数学, 1982(2): 45–54.
WANG Peizhuang, SUGENO M. The factors field and background structure for fuzzy subsets[J]. Fuzzy mathematics, 1982(2): 45–54.
- [2] 汪培庄, 李洪兴. 知识表示的数学理论 [M]. 天津: 天津科学技术出版社, 1994.
- [3] 李洪兴. 因素空间与模糊决策 [J]. 北京师范大学学报(自然科学版), 1994, 30(1): 41–46.
LI Hongxing. Factor spaces and fuzzy decision making[J]. Journal of Beijing Normal University (natural science edition), 1994, 30(1): 41–46.
- [4] 袁学海, 汪培庄. 因素空间和范畴 [J]. 模糊系统与数学, 1995, 9(2): 25–33.
YUAN Xuehai, WANG Peizhuang. Factor spaces and categories[J]. Fuzzy systems and mathematics, 1995, 9(2): 25–33.
- [5] WANG Peizhuang, LIU Zengliang, SHI Yong, et al. Factor space, the theoretical base of data science[J]. *Annals of data science*, 2014, 1(2): 233–251.
- [6] 汪培庄, 郭嗣琮, 包研科, 等. 因素空间中的因素分析法 [J]. 辽宁工程技术大学学报(自然科学版), 2014, 33(7): 865–870.
WANG Peizhuang, GUO Sicong, BAO Yanke, et al. Causality analysis in factor spaces[J]. Journal of Liaoning Technical University (natural science edition), 2014, 33(7): 865–870.
- [7] 于福生, 罗承忠. 粒子因素空间与智能诊断专家系统 [C]// 第七届全国电工数学学术年会论文集. 北京: 科学技术出版社, 1999: 24–27.
YU Fusheng, LUO Chengzhong. Granule factors space and intelligent diagnostic expert systems[C]//Proceedings of the 7th National Conference on Electric Mathematics. Beijing: China Science & Technology Press, 1999: 24–27.
- [8] 刘海涛, 郭嗣琮. 因素分析法的推理模型 [J]. 辽宁工程技术大学学报(自然科学版), 2015, 34(1): 124–128.
LIU Haitao, GUO Sicong. Reasoning model of causality analysis[J]. Journal of Liaoning Technical University (natural science edition), 2015, 34(1): 124–128.
- [9] 曾繁慧, 李艺. 因素空间理论的决策树 C4.5 算法改进 [J]. 辽宁工程技术大学学报(自然科学版), 2017, 36(1): 109–112.
ZENG Fanhui, LI Yi. An improved decision tree algorithm based on factor space theory[J]. Journal of Liaoning Technical University (natural science edition), 2017, 36(1): 109–112.
- [10] 汪华东, 郭嗣琮. 基于因素空间反馈外延外包络的 DFE 决策 [J]. 计算机工程与应用, 2015, 51(15): 148–152, 156.
WANG Huadong, GUO Sicong. DFE decision-making based on outer-enveloping of feedback extension in factor spaces[J]. Computer engineering and applications, 2015, 51(15): 148–152, 156.
- [11] 孙慧, 曾繁慧, 蒲凌杰. 因素空间理论下多目标因果分析的降维算法 [J]. 辽宁工程技术大学学报(自然科学版), 2021, 40(5): 466–472.
SUN Hui, ZENG Fanhui, PU Lingjie. Research on dimension reduction algorithm of multiple targets causal analysis based on factor space theory[J]. Journal of Liaoning Technical University (natural science edition), 2021, 40(5): 466–472.
- [12] HELLSTRÖM T. The relevance of causation in robotics: a review, categorization, and analysis[J]. *Paladyn, journal of behavioral robotics*, 2021, 12(1): 238–255.
- [13] 蒲凌杰, 曾繁慧, 汪培庄. 因素空间理论下基点分类算法研究 [J]. 智能系统学报, 2020, 15(3): 528–536.
PU Lingjie, ZENG Fanhui, WANG Peizhuang. Base point classification algorithm based on factor space theory[J]. CAAI transactions on intelligent systems, 2020, 15(3): 528–536.
- [14] 刘海涛, 郝传波, 傅贵. 因素空间下的煤与瓦斯突出预测方法 [J]. 黑龙江科技大学学报, 2017, 27(4): 354–358.
LIU Haitao, HAO Chuanbo, FU Gui. Study on factor space-based prediction method of coal and gas outburst[J]. Journal of Heilongjiang University of Science & Technology, 2017, 27(4): 354–358.
- [15] REICHENBACH H, REICHENBACH M. The direction of time[M]. Berkeley: University of California Press, 1956.
- [16] HÁJEK A, HITCHCOCK C. The Oxford handbook of probability and philosophy[M]. New York: Oxford University Press, 2016.
- [17] SPRENGER J. Foundations of a probabilistic theory of causal strength[J]. *The philosophical review*, 2018,

127(3): 371–398.

- [18] FAGHIHI U, BOUCHARD S M, BISKRI I. Science of data: a new ladder for causation[M]//Sayed-Mouchaweh M. Explainable AI Within the Digital Transformation and Cyber Physical Systems. Cham: Springer, 2021: 33–45.

- [19] 黄影平. 贝叶斯网络发展及其应用综述 [J]. 北京理工大学学报, 2013, 33(12): 1211–1219.

HUANG Yingping. Survey on Bayesian network development and application[J]. Transactions of Beijing Institute of Technology, 2013, 33(12): 1211–1219.

- [20] 曲国华, 曾繁慧, 刘增良, 等. 因素空间中的背景分布与模糊背景关系 [J]. 模糊系统与数学, 2017, 31(6): 66–73.

QU Guohua, ZENG Fanhui, LIU Zengliang, et al. Background distribution and fuzzy background relation in factor spaces[J]. Fuzzy systems and mathematics, 2017, 31(6): 66–73.

- [21] 张艳妮, 曾繁慧, 郭嗣琮. 因素空间理论的 Markov 过程 [J]. 辽宁工程技术大学学报(自然科学版), 2019, 38(4): 385–389.

ZHANG Yanni, ZENG Fanhui, GUO Sicong. Factor Markov process of factor space theory[J]. Journal of Liaoning Technical University (natural science edition), 2019, 38(4): 385–389.

作者简介:



曾繁慧, 教授, 主要研究方向为基于因素空间的数据挖掘理论与应用。入选“辽宁省兴辽英才计划”, 主持、参与完成中国工程院重点项目、国家自然科学基金项目、辽宁省科研基金项目等 10 余项。获首批国家一流课程、辽宁省研究生教学成果一等奖等 80 余项奖励。发表学术论文 70 余篇。E-mail: 597873883@qq.com。



胡光闪, 硕士研究生, 主要研究方向为因素空间理论下的数据挖掘和智能决策。E-mail: 1599546002@qq.com。



汪培庄, 教授, 博士生导师, 主要研究方向为模糊数学及其在人工智能中的应用, 是我国模糊数学的传播者和主要学术带头人, 曾任国际模糊系统协会副主席, 于 1982 年在国际上创立因素空间的数学理论, 现被公认为人工智能的数学基础理论。现为辽宁工程技术大学的特聘教授, 并任智能工程与数学研究院院长, 荣获中华人民共和国 70 周年纪念章。发表学术论文 200 余篇, 出版学术著作 10 余部。E-mail: peizhuangw@126.com。

2024 中国元宇宙大会

随着虚拟现实、数字孪生、人工智能、人机交互等前沿科技的融合发展, 元宇宙已从科幻想象走向现实, 成为一个应用前景和商业价值不可限量的新产业赛道。

为促进产学研成果交流合作, 推动元宇宙相关技术创新及元宇宙产业生态健康发展, 7 月 13—14 日, 由中国人工智能学会主办的 2024 中国元宇宙大会将在杭州举办。

大会将组织进行多场主题报告, 举办一系列专题论坛, 邀请元宇宙领域的科技企业、学术机构的行业专家和技术精英共同分享研究成果, 展示技术方案, 深入探讨元宇宙的新场景、新应用、新生态。

元宇宙技术的不断发展, 其应用场景将越来越广泛, 深刻改变人们的工作、生活和娱乐方式。2024 中国元宇宙大会不仅是元宇宙技术的一次深度交流, 更是未来科技形态的一次前瞻讨论, 将引领我们跨越现实与虚拟的鸿沟, 走进精彩无限的元宇宙世界。